



Europäisches Patentamt  
European Patent Office  
Office européen des brevets



(11) Publication number : **0 399 666 B1**

(12)

## EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification :  
**28.07.93 Bulletin 93/30**

(51) Int. Cl.<sup>5</sup> : **C12N 15/62, C07K 13/00,  
C12P 21/02**

(21) Application number : **90304575.5**

(22) Date of filing : **26.04.90**

(54) **Fusion proteins containing N-terminal fragments of human serum albumin.**

Consolidated with 90907285.2/0470165  
(European application No./publication No.) by  
decision dated 20.07.92.

(30) Priority : **29.04.89 GB 8909919**

(43) Date of publication of application :  
**28.11.90 Bulletin 90/48**

(45) Publication of the grant of the patent :  
**28.07.93 Bulletin 93/30**

(84) Designated Contracting States :  
**AT BE CH DE DK ES FR GR IT LI LU NL SE**

(56) References cited :  
**EP-A- 0 308 381  
EP-A- 0 322 094**

(73) Proprietor : **Delta Biotechnology Limited  
Castle Court, Castle Boulevard  
Nottingham NG7 1FD (GB)**

(72) Inventor : **Ballance, David James  
11 South Road  
West Bridgford, Nottingham NG2 7AG (GB)**

(74) Representative : **Bassett, Richard Simon et al  
ERIC POTTER & CLARKSON St. Mary's Court  
St. Mary's Gate  
Nottingham NG1 1LE (GB)**

**EP 0 399 666 B1**

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid (Art. 99(1) European patent convention).

## Description

The present invention relates to fusion polypeptides where two individual polypeptides or parts thereof are fused to form a single amino acid chain. Such fusion may arise from the expression of a single continuous coding sequence formed by recombinant DNA techniques.

Fusion polypeptides are known, for example those where a polypeptide which is the ultimately desired product of the process is expressed with an N-terminal "leader sequence" which encourages or allows secretion of the polypeptide from the cell. An example is disclosed in EP-A-116 201 (Chiron).

Human serum albumin (HSA) is a known protein found in the blood. EP-A-147 198 (Delta Biotechnology) discloses its expression in a transformed host, in this case yeast. Our earlier application EP-A-322 094 discloses N-terminal fragments of HSA, namely those consisting of residues 1-n where n is 369 to 419, which have therapeutic utility. The application also mentions the possibility of fusing the C-terminal residue of such molecules to other, unnamed, polypeptides.

One aspect of the present invention provides a fusion polypeptide comprising, as at least part of the N-terminal portion thereof, an N-terminal portion of HSA or a variant thereof and, as at least part of the C-terminal portion thereof, another polypeptide except that, when the said N-terminal portion of HSA is the 1-n portion where n is 369 to 419 or a variant thereof then the said polypeptide is (a) the 585 to 1578 portion of human fibronectin or a variant thereof, (b) the 1 to 368 portion of CD4 or a variant thereof, (c) platelet derived growth factor, or a variant thereof, (d) transforming growth factor, or a variant thereof, (e) the 1-261 portion of mature human plasma fibronectin or a variant thereof, (f) the 278-578 portion of mature human plasma fibronectin or a variant thereof, (g) the 1-272 portion of mature human von Willebrand's Factor or a variant thereof, or (h) alpha-1-antitrypsin or a variant thereof.

The N-terminal portion of HSA is preferably the said 1-n portion, the 1-177 portion (up to and including the cysteine), the 1-200 portion (up to but excluding the cysteine) or a portion intermediate 1-177 and 1-200.

The term "human serum albumin" (HSA) is intended to include (but not necessarily to be restricted to) known or yet-to-be-discovered polymorphic forms of HSA. For example, albumin Naskapi has Lys-372 in place of Glu-372 and pro-albumin Christchurch has an altered pro-sequence. The term "variants" is intended to include (but not necessarily to be restricted to) minor artificial variations in sequence (such as molecules lacking one or a few residues, having conservative substitutions or minor insertions of residues, or having minor variations of amino acid structure). Thus polypeptides which have 80%, preferably 85%, 90%, 95% or 99%, homology with HSA are deemed to be "variants". It is also preferred for such variants to be physiologically equivalent to HSA; that is to say, variants preferably share at least one pharmacological utility with HSA. Furthermore, any putative variant which is to be used pharmacologically should be non-immunogenic in the animal (especially human) being treated.

Conservative substitutions are those where one or more amino acids are substituted for others having similar properties such that one skilled in the art of polypeptide chemistry would expect at least the secondary structure, and preferably the tertiary structure, of the polypeptide to be substantially unchanged. For example, typical such substitutions include asparagine for glutamine, serine for asparagine and arginine for lysine. Variants may alternatively, or as well, lack up to ten (preferably only one or two) intermediate amino acid residues (ie not at the termini of the said N-terminal portion of HSA) in comparison with the corresponding portion of natural HSA; preferably any such omissions occur in the 100 to 369 portion of the molecule (relative to mature HSA itself) (if present). Similarly, up to ten, but preferably only one or two, amino acids may be added, again in the 100 to 369 portion for preference (if present). The term "physiologically functional equivalents" also encompasses larger molecules comprising the said sequence plus a further sequence at the N-terminal (for example, pro-HSA, pre-pro-HSA and met-HSA).

Clearly, the said "another polypeptide" in the fusion compounds of the invention cannot be the remaining portion of HSA, since otherwise the whole polypeptide would be HSA, which would not then be a "fusion polypeptide".

Even when the HSA-like portion is not the said 1-n portion of HSA, it is preferred for the non-HSA portion to be one of the said (a) to (h) entities.

The 1 to 368 portion of CD4 represents the first four disulphide-linked immunoglobulin-like domains of the human T lymphocyte CD4 protein, the gene for and amino acid sequence of which are disclosed in D. Smith *et al* (1987) Science 328, 1704-1707. It is used to combat HIV infections.

The sequence of human platelet-derived growth factor (PDGF) is described in Collins *et al* (1985) Nature 316, 748-750. Similarly, the sequence of transforming growth factors  $\beta$  (TGF- $\beta$ ) is described in Derynck *et al* (1985) Nature 316, 701-705. These growth factors are useful for wound-healing.

A cDNA sequence for the 1-261 portion of Fn was disclosed in EP-A-207 751 (obtained from plasmid pFH6 with endonuclease PvuII). This portion binds fibrin and can be used to direct fused compounds to blood clots.

A cDNA sequence for the 278-578 portion of Fn, which contains a collagen-binding domain, was disclosed by R.J. Owens and F.E. Baralle in 1986 E.M.B.O.J. 5, 2825-2830. This portion will bind to platelets.

The 1-272 portion of von Willebrand's Factor binds and stabilises factor VIII. The sequence is given in Bontham *et al*, Nucl. Acids Res. 14, 7125-7127.

5 Variants of alpha-1-antitrypsin include those disclosed by Rosenberg *et al* (1984) Nature 312, 77-80. In particular, the present invention includes the Pittsburgh variant (Met<sup>358</sup> is mutated to Arg) and the variant where Pro<sup>357</sup> and Met<sup>358</sup> are mutated to alanine and arginine respectively. These compounds are useful in the treatment of septic shock and lung disorders.

10 Variants of the non-HSA portion of the polypeptides of the invention include variations as discussed above in relation to the HSA portion, including those with conservative amino acid substitutions, and also homologues from other species.

The fusion polypeptides of the invention may have N-terminal amino acids which extend beyond the portion corresponding to the N-terminal portion of HSA. For example, if the HSA-like portion corresponds to an N-terminal portion of mature HSA, then pre-, pro-, or pre-pro sequences may be added thereto, for example 15 the yeast alpha-factor leader sequence. The fused leader portions of WO 90/01063 may be used. The polypeptide which is fused to the HSA portion may be a naturally-occurring polypeptide, a fragment thereof or a novel polypeptide, including a fusion polypeptide. For example, in Example 3 below, a fragment of fibronectin is fused to the HSA portion via a 4 amino acid linker.

20 It has been found that the amino terminal portion of the HSA molecule is so structured as to favour particularly efficient translocation and export of the fusion compounds of the invention in eukaryotic cells.

A second aspect of the invention provides a transformed host having a nucleotide sequence so arranged as to express a fusion polypeptide as described above. By "so arranged", we mean, for example, that the nucleotide sequence is in correct reading frame with an appropriate RNA polymerase binding site and translation start sequence and is under the control of a suitable promoter. The promoter may be homologous with or heterologous to the host. Downstream (3') regulatory sequences may be included if desired, as is known. The 25 host is preferably yeast (for example *Saccharomyces* spp., e.g. *S. cerevisiae*; *Kluyveromyces* spp., e.g. *K. lactis*; *Pichia* spp.; or *Schizosaccharomyces* spp., e.g. *S. pombe*) but may be any other suitable host such as *E. coli*, *B. subtilis*, *Aspergillus* spp., mammalian cells, plant cells or insect cells.

30 A third aspect of the invention provides a process for preparing a fusion polypeptide according to the first aspect of the invention by cultivation of a transformed host according to the second aspect of the invention, followed by separation of the fusion polypeptide in a useful form.

A fourth aspect of the invention provides therapeutic methods of treatment of the human or other animal body comprising administration of such a fusion polypeptide.

35 In the methods of the invention we are particularly concerned to improve the efficiency of secretion of useful therapeutic human proteins from yeast and have conceived the idea of fusing to amino-terminal portions of HSA those proteins which may ordinarily be only inefficiently secreted. One such protein is a potentially valuable wound-healing polypeptide representing amino acids 585 to 1578 of human fibronectin (referred to herein as Fn 585-1578). As we have described in a separate application (filed simultaneously herewith) this molecule contains cell spreading, chemotactic and chemokinetic activities useful in healing wounds. The fusion 40 polypeptides of the present invention wherein the C-terminal portion is Fn 585-1578 can be used for wound healing applications as biosynthesised, especially where the hybrid human protein will be topically applied. However, the portion representing amino acids 585 to 1578 of human fibronectin can if desired be recovered from the fusion protein by preceding the first amino acid of the fibronectin portion by amino acids comprising a factor X cleavage site. After isolation of the fusion protein from culture supernatant, the desired molecule is 45 released by factor X cleavage and purified by suitable chromatography (e.g. ion-exchange chromatography). Other sites providing for enzymatic or chemical cleavage can be provided, either by appropriate juxtaposition of the N-terminal and C-terminal portions or by the insertion therebetween of an appropriate linker.

50 At least some of the fusion polypeptides of the invention, especially those including the said CD4 and vWF fragments, PDGF and  $\alpha_1$ AT, also have an increased half-life in the blood and therefore have advantages and therapeutic utilities themselves, namely the therapeutic utility of the non-HSA portion of the molecule. In the case of  $\alpha_1$ AT and others, the compound will normally be administered as a one-off dose or only a few doses over a short period, rather than over a long period, and therefore the compounds are less likely to cause an immune response.

## 55 EXAMPLES : SUMMARY

Standard recombinant DNA procedures were as described by Maniatis *et al* (1982 and recent 2nd edition) unless otherwise stated. Construction and analysis of phage M13 recombinant clones was as described by

Messing (1983) and Sanger *et al* (1977).

DNA sequences encoding portions of human serum albumin used in the construction of the following molecules are derived from the plasmids mHOB12 and pDBD2 (EP-A-322 094, Delta Biotechnology Ltd, relevant portions of which are reproduced below) or by synthesis of oligonucleotides equivalent to parts of this sequence. DNA sequences encoding portions of human fibronectin are derived from the plasmid pFHDEL1, or by synthesis of oligonucleotides equivalent to parts of this sequence. Plasmid pFHDEL1, which contains the complete human cDNA encoding plasma fibronectin, was obtained by ligation of DNA derived from plasmids pFH6, 16, 54, 154 and 1 (EP-A-207 751; Delta Biotechnology Ltd).

This DNA represents an mRNA variant which does not contain the 'ED' sequence and had an 89-amino acid variant of the III-CS region (R.J. Owens, A.R. Kornblihtt and F.E. Baralle (1986) Oxford Surveys on Eukaryotic Genes 3 141-160). The map of this vector is disclosed in Fig. 11 and the protein sequence of the mature polypeptide produced by expression of this cDNA is shown in Fig. 5.

Oligonucleotides were synthesised on an Applied Biosystems 380B oligonucleotide synthesiser according to the manufacturer's recommendations (Applied Biosystems, Warrington, Cheshire, UK).

An expression vector was constructed in which DNA encoding the HSA secretion signal and mature HSA up to and including the 387th amino acid, leucine, fused in frame to DNA encoding a segment of human fibronectin representing amino acids 585 to 1578 inclusive, was placed downstream of the hybrid promoter of EP-A-258 067 (Delta Biotechnology), which is a highly efficient galactose-inducible promoter functional in *Saccharomyces cerevisiae*. The codon for the 1578th amino acid of human fibronectin was directly followed by a stop codon (TAA) and then the *S. cerevisiae* phosphoglycerate kinase (PGK) gene transcription terminator. This vector was then introduced into *S. cerevisiae* by transformation, wherein it directed the expression and secretion from the cells of a hybrid molecule representing the N-terminal 387 amino acids of HSA C-terminally fused to amino acids 585 to 1578 of human fibronectin.

In a second example a similar vector is constructed so as to enable secretion by *S. cerevisiae* of a hybrid molecule representing the N-terminal 195 amino acids of HSA C-terminally fused to amino acids 585 to 1578 of human fibronectin.

Aspects of the present invention will now be described by way of example and with reference to the accompanying drawings, in which:

Figure 1 (on two sheets) depicts the amino acid sequence currently thought to be the most representative of natural HSA, with (boxed) the alternative C-termini of HSA(1-n);

Figure 2 (on two sheets) depicts the DNA sequence coding for mature HSA, wherein the sequence included in Linker 3 is underlined;

Figure 3 illustrates, diagrammatically, the construction of mHOB16;

Figure 4 illustrates, diagrammatically, the construction of pHOB31;

Figure 5 (on 6 sheets) illustrates the mature protein sequence encoded by the Fn plasmid pFHDEL1;

Figure 6 illustrates Linker 5, showing the eight constituent oligonucleotides;

Figure 7 shows schematically the construction of plasmid pDBDF2;

Figure 8 shows schematically the construction of plasmid pDBDF5;

Figure 9 shows schematically the construction of plasmid pDBDF9;

Figure 10 shows schematically the construction of plasmid DBDF12, using plasmid pFHDEL1; and

Figure 11 shows a map of plasmid pFHDEL1.

#### EXAMPLE 1 : HSA 1-387 FUSED TO Fn 585-1578

The following is an account of a preparation of plasmids comprising sequences encoding a portion of HSA, as is disclosed in EP-A-322 094.

The human serum albumin coding sequence used in the construction of the following molecules is derived from the plasmid M13mp19.7 (EP-A-201 239, Delta Biotechnology Ltd.) or by synthesis of oligonucleotides equivalent to parts of this sequence. Oligonucleotides were synthesised using phosphoramidite chemistry on an Applied Biosystems 380B oligonucleotide synthesizer according to the manufacturer's recommendations (AB Inc., Warrington, Cheshire, England).

An oligonucleotide was synthesised (Linker A) which represented a part of the known HSA coding sequence (Figure 2) from the *Pst*I site (1235-1240, Figure 2) to the codon for valine 381 wherein that codon was changed from GTG to GTC:

Linker 1 was ligated into the vector M13mp19 (Norranders *et al.*, 1983) which had been digested with *Pst*I and *Hinc*II and the ligation mixture was used to transfect *E. coli* strain XL1-Blue (Stratagene Cloning Systems, San Diego, CA). Recombinant clones were identified by their failure to evolve a blue colour on medium containing the chromogenic indicator X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside) in the presence of IPTG (isopropylthio- $\beta$ -galactoside). DNA sequence analysis of template DNA prepared from bacteriophage particles of recombinant clones identified a molecule with the required DNA sequence, designated mHOB12 (Figure 3).

35

45 (EP-A-210 239). M13mp19.7 was digested with XhoI and made flush-ended by S1-nuclease treatment and was then ligated with the following oligonucleotide (Linker 2):

50

55

5

plaques and then analysed by DNA sequencing to identify a clone, pDBD1 (Figure 4), with the correct sequence.

A 1.1 kb HindIII to PstI fragment representing the 5' end of the HSA coding region and one half of the inserted oligonucleotide linker was isolated from pDBD1 by agarose gel electrophoresis. This fragment was then ligated with double stranded mHOB12 previously digested with HindIII and PstI and the ligation mix was then used to transfect E.coli XL1-Blue. Single stranded template DNA was prepared from mature bacteriophage particles of several plaques. The DNA was made double stranded in vitro by extension from annealed sequencing primer with the Klenow fragment of DNA polymerase I in the presence of deoxynucleoside triphosphates. Restriction enzyme analysis of this DNA permitted the identification of a clone with the correct configuration, mHOB15 (Figure 4).

The following oligonucleotide (Linker 3) represents from the codon for the 382nd amino acid of mature HSA (glutamate, GAA) to the codon for lysine 389 which is followed by a stop codon (TAA) and a HindIII site and then a BamHI cohesive end:

#### Linker 3

E E P Q N L I K J

5' GAA GAG CCT CAG AAT TTA ATC AAA TAA GCTTG 3'

3' CTT CTC GGA GTC TTA AAT TAG TTT ATT CGAACCTAG 5'

This was ligated into double stranded mHOB15, previously digested with HincII and BamHI. After ligation, the DNA was digested with HincII to destroy all non-recombinant molecules and then used to transfect E.coli XL1-Blue. Single stranded DNA was prepared from bacteriophage particles of a number of clones and subjected to DNA sequence analysis. One clone having the correct DNA sequence was designated mHOB16 (Figure 4).

A molecule in which the mature HSA coding region was fused to the HSA secretion signal was created by insertion of Linker 4 into BamHI and XhoI digested M13mp19.7 to form pDBD2 (Figure 4).

#### Linker 4

M K W V S F

5' GATCC ATG AAG TGG GTA AGC TTT

G TAC TTC ACC CAT TCG AAA

I S L L F L F S

ATT TCC CTT CTT TTT CTC TTT AGC

TAA AGG GAA GAA AAA GAG AAA TCG

	S	A	Y	S	R	G	V	F
	TCG	GCT	TAT	TCC	AGG	GGT	GTG	TTT
5	AGC	CGA	ATA	AGG	TCC	CCA	CAC	AAA
10	R	R						
	CG	3'						
	GCAGCT	5'						

15

In this linker the codon for the fourth amino acid after the initial methionine, ACC for threonine in the HSA pre-pro leader sequence (Lawn *et al.*, 1981), has been changed to AGC for serine to create a *Hind*III site.

A sequence of synthetic DNA representing a part of the known HSA coding sequence (Lawn *et al.*, 1981) (amino acids 382 to 387, Fig. 2), fused to part of the known fibronectin coding sequence (Kornblihtt *et al.*, 1985) (amino acids 585 to 640, Fig. 2), was prepared by synthesising six oligonucleotides (Linker 5, Fig. 6). The oligonucleotides 2, 3, 4, 6, 7 and 8 were phosphorylated using T4 polynucleotide kinase and then the oligonucleotides were annealed under standard conditions in pairs, i.e. 1+8, 2+7, 3+6 and 4+5. The annealed oligonucleotides were then mixed together and ligated with mHOB12 which had previously been digested with the restriction enzymes *Hinc*II and *Eco*RI. The ligation mixture was then used to transfect *E.coli* XL1-Blue (Stratagene Cloning Systems, San Diego, CA). Single stranded template DNA was then prepared from mature bacteriophage particles derived from several independent plaques and then was analysed by DNA sequencing. A clone in which a linker of the expected sequence had been correctly inserted into the vector was designated pDBDF1 (Fig. 7). This plasmid was then digested with *Pst*I and *Eco*RI and the approx. 0.24kb fragment was purified and then ligated with the 1.29kb *Bam*HI-*Pst*I fragment of pDBD2 (Fig. 7) and *Bam*HI + *Eco*RI digested pUC19 (Yanisch-Perron, *et al.*, 1985) to form pDBDF2 (Fig. 7).

A plasmid containing a DNA sequence encoding full length human fibronectin, pFHDEL1, was digested with *Eco*RI and *Xho*I and a 0.77kb *Eco*RI-*xho*I fragment (Fig. 8) was isolated and then ligated with *Eco*RI and *sall* digested M13 mp18 (Norrander *et al.*, 1983) to form pDBDF3 (Fig. 8).

The following oligonucleotide linker (Linker 6) was synthesised, representing from the *Pst*I site at 4784-4791 of the fibronectin sequence of EP-A-207 751 to the codon for tyrosine 1578 (Fig. 5) which is followed by a stop codon (TAA), a *Hind*III site and then a *Bam*HI cohesive end:

#### Linker 6

40

	G	P	D	Q	T	E	M	T	I	E	G	L
	GGT	CCA	GAT	CAA	ACA	GAA	ATG	ACT	ATT	GAA	GGC	TTG
45	A	CGT	CCA	GGT	CTA	GTT	TGT	CTT	TAC	TGA	TAA	CTT
											CCG	AAC
50	Q	P	T	V	E	Y		Stop				
	CAG	CCC	ACA	GTG	GAG	TAT		TAA		GCTTG		
	GTC	GGG	TGT	CAC	CTC	ATA		ATT		CGAACCTAG		

55

This linker was then ligated with *Pst*I and *Hind*III digested pDBDF3 to form pDBDF4 (Fig. 8). The following DNA fragments were then ligated together with *Bgl*II digested pKV50 (EP-A-258 067) as shown in Fig. 8: 0.68kb *Eco*RI-*Bam*HI fragment of pDBDF4, 1.5kb *Bam*HI-*Stu*I fragment of pDBDF2 and the 2.2kb *Stu*I-*Eco*RI fragment of pFHDEL1. The resultant plasmid pDBDF5 (Fig. 8) includes the promoter of EP-A-258 067 to direct the ex-

pression of the HSA secretion signal fused to DNA encoding amino acids 1-387 of mature HSA, in turn fused directly and in frame with DNA encoding amino acids 585-1578 of human fibronectin, after which translation would terminate at the stop codon TAA. This is then followed by the *S.cerevisiae* PGK gene transcription terminator. The plasmid also contains sequences which permit selection and maintenance in *Escherichia coli* and *S.cerevisiae* (EP-A-258 067).

This plasmid was introduced into *S.cerevisiae* S150-2B (*leu2-3 leu2-112 ura3-52 trp1-289 his3-1*) by standard procedures (Beggs, 1978). Transformants were subsequently analysed and found to produce the HSA-fibronectin fusion protein.

#### 10 EXAMPLE 2 : HSA 1-195 FUSED TO Fn 585-1578

In this second example the first domain of human serum albumin (amino acids 1-195) is fused to amino acids 585-1578 of human fibronectin.

The plasmid pDBD2 was digested with *Bam*HI and *Bgl*II and the 0.79kb fragment was purified and then  
15 ligated with *Bam*HI-digested M13mp19 to form pDBDF6 (Fig. 6). The following oligonucleotide:

5'-C C A A A G C T C G A G G A A C T T C G-3'

was used as a mutagenic primer to create a *Xho*I site in pDBDF6 by *in vitro* mutagenesis using a kit supplied  
20 by Amersham International PLC. This site was created by changing base number 696 of HSA from a T to a G (Fig. 2). The plasmid thus formed was designated pDBDF7 (Fig. 9). The following linker was then synthesised to represent from this newly created *Xho*I site to the codon for lysine 195 of HSA (AAA) and then from the codon for isoleucine 585 of fibronectin to the ends of oligonucleotides 1 and 8 shown in Fig. 6.

#### 25 Linker 7

                  D   E   L   R   D   E   G   K   A   S   S   A   K  
30   TC GAT GAA CTT CGG GAT GAA GGG AAG GCT TCG TCT GCC AAA  
                  A   CTT GAA GCC CTA CTT CCC TTC CGA AGC AGA CGG TTT  
  
                  I   T   E   T   P   S   Q   P   N   S   H  
                  ATC ACT GAG ACT CCG AGT CAG C  
40   TAG TGA CTC TGA GGC TCA GTC GGG TTG AGG GTG G

This linker was ligated with the annealed oligonucleotides shown in Fig. 3, i.e. 2+7, 3+6 and 4+5 together with *Xho*I and *Eco*RI digested pDBDF7 to form pDBDF8 (Fig. 9). Note that in order to recreate the original HSA DNA sequence, and hence amino acid sequence, insertion of linker 7 and the other oligonucleotides into  
45 pDBDF7 does not recreate the *Xho*I site.

The 0.83kb *Bam*HI-*Stu*I fragment of pDBDF8 was purified and then was ligated with the 0.68kb *Eco*RI-*Bam*HI fragment of pDBDF2 and the 2.22kb *Stu*I-*Eco*RI fragment of pFHDEL1 into *Bgl*II-digested pKV50 to form pDBDF9 (Fig. 9). This plasmid is similar to pDBDF5 except that it specifies only residues 1-195 of HSA rather than 1-387 as in pDBDF5.

50 When introduced into *S.cerevisiae* S150-2B as above, the plasmid directed the expression and secretion of a hybrid molecule composed of residues 1-195 of HSA fused to residues 585-1578 of fibronectin.

#### EXAMPLE 3 : HSA 1-387 FUSED TO Fn 585-1578, AS CLEAVABLE MOLECULE

55 In order to facilitate production of large amounts of residues 585-1578 of fibronectin, a construct was made in which DNA encoding residues 1-387 of HSA was separated from DNA encoding residues 585-1578 of fibronectin by the sequence

I E G R  
 ATT GAA GGT AGA  
 TAA CTT CCA TCT

5

which specifies the cleavage recognition site for the blood clotting Factor X. Consequently the purified secreted product can be treated with Factor X and then the fibronectin part of the molecule can be separated from the HSA part.

10

To do this two oligonucleotides were synthesised and then annealed to form Linker 8.

#### Linker 8

15

E E P Q N L I E G  
 GAA GAG CCT CAG AAT TTA ATT GAA GGT  
 CTT CTC GGA GTC TTA AAT TAA CTT CCA

20

R I T E T P S Q P  
 AGA ATC ACT GAG ACT CCG AGT CAG C  
 TCT TAG TGA CTC TGA GGC TCA GTC GGG

25

30

N S H

35

TTG AGG GTG G

This linker was then ligated with the annealed oligonucleotides shown in Fig. 6, i.e. 2+7, 3+6 and 4+5 into HincII and EcoRI digested mHOB12, to form pDBDF10 (Fig. 7). The plasmid was then digested with PstI and EcoRI and the roughly 0.24kb fragment was purified and then ligated with the 1.29kb BamHI-PstI fragment of pDBD2 and BamHI and EcoRI digested pUC19 to form pDBDF11 (Fig. 10).

40

The 1.5kb BamHI-StuI fragment of pDBDF11 was then ligated with the 0.68kb EcoRI-BamHI fragment of pDBDF4 and the 2.22kb StuI-EcoRI fragment of pFHDEL1 into BglII-digested pKV50 to form pDBDF12 (Fig. 10). This plasmid was then introduced into *S.cerevisiae* S150-2B. The purified secreted fusion protein was treated with Factor X to liberate the fibronectin fragment representing residues 585-1578 of the native molecule.

45

#### REFERENCES

- Beggs, J.D. (1978) *Nature* **275**, 104-109  
 Kornblihtt *et al.* (1985) *EMBO J.* **4**, 1755-1759  
 Lawn, R.M. *et al.* (1981) *Nucl. Acid. Res.* **9**, 6103-6114  
 Maniatis, T. *et al.* (1982) *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.  
 Messing, J. (1983) *Methods Enzymol.* **101**, 20-78  
 Norrander, J. *et al.* (1983) *Gene* **26**, 101-106  
 Sanger, F. *et al.* (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467  
 Yanisch-Perron, C. (1985) *Gene* **33**, 103-119

55

**Claims****Claims for the following Contracting States : AT, BE, CH, LI, DE, DK, FR, IT, LU, NL, SE**

- 5 1. A fusion polypeptide comprising, as at least part of the N-terminal portion thereof, an N-terminal portion of HSA or a variant thereof and, as at least part of the C-terminal portion thereof, another polypeptide except that, when the said N-terminal portion of HSA is the 1-n portion where n is 369 to 419 or a variant thereof then the said polypeptide is (a) the 585 to 1578 portion of human fibronectin or a variant thereof, (b) the 1 to 368 portion of CD4 or a variant thereof, (c) platelet derived growth factor or a variant thereof, (d) transforming growth factor  $\beta$  or a variant thereof, (e) the 1-261 portion of mature human plasma fibronectin or a variant thereof, (f) the 278-578 portion of mature human plasma fibronectin or a variant thereof, (g) the 1-272 portion of mature human von Willebrand's Factor or a variant thereof, or (h) alpha-1-antitrypsin or a variant thereof.
- 10 2. A fusion polypeptide according to Claim 1 additionally comprising at least one N-terminal amino acid extending beyond the portion corresponding to the N-terminal portion of HSA.
3. A fusion polypeptide according to Claim 1 or 2 wherein there is a cleavable region at the junction of the said N-terminal or C-terminal portions.
- 20 4. A fusion polypeptide according to any one of the preceding claims wherein the said C-terminal portion is the 585 to 1578 portion of human plasma fibronectin or a variant thereof.
5. A transformed or transfected host having a nucleotide sequence so arranged as to express a fusion polypeptide according to any one of the preceding claims.
- 25 6. A process for preparing a fusion polypeptide by cultivation of a host according to Claim 5, followed by separation of the fusion polypeptide in a useful form.
7. A fusion polypeptide according to any one of Claims 1 to 4 for use in therapy.

30  
**Claims for the following Contracting States : ES, GR**

- 35 1. A process for preparing a fusion polypeptide by (i) cultivation of a transformed or transfected host having a nucleotide sequence so arranged as to express a fusion polypeptide, followed by (ii) separation of the fusion polypeptide in a useful form, characterised in that the fusion polypeptide comprises as at least part of the N-terminal portion thereof, an N-terminal portion of HSA or a variant thereof and, as at least part of the C-terminal portion thereof, another polypeptide except that, when the said N-terminal portion of HSA is the 1-n portion where n is 369 to 419 or a variant thereof then the said polypeptide is (a) the 585 to 1578 portion of human fibronectin or a variant thereof, (b) the 1 to 368 portion of CD4 or a variant thereof, (c) platelet derived growth factor or a variant thereof, (d) transforming growth factor  $\beta$  or a variant thereof, (e) the 1-261 portion of mature human plasma fibronectin or a variant thereof, (f) the 278-578 portion of mature human plasma fibronectin or a variant thereof, (g) the 1-272 portion of mature human von Willebrand's Factor or a variant thereof, or (h) alpha-1-antitrypsin or a variant thereof.
- 40 2. A process according to Claim 1, wherein the fusion polypeptide additionally comprising at least one N-terminal amino acid extending beyond the portion corresponding to the N-terminal portion of HSA.
3. A process according to Claim 1 or 2 wherein, in the fusion polypeptide, there is a cleavable region at the junction of the said N-terminal or C-terminal portions.
- 50 4. A process according to any one of the preceding claims wherein the said C-terminal portion is the 585 to 1578 portion of human plasma fibronectin or a variant thereof.

55

**Patentansprüche****Patentansprüche für folgende Vertragsstaaten : AT, BE, CH, DE, DK, FR, IT, LU, NL, SE**

- 5 1. Fusionspolypeptid, umfassend als mindestens einen Teil seines N-terminalen Teils einen N-terminalen Teil von HSA oder eine Variante davon und als mindestens einen Teil seines C-terminalen Teils ein weiteres Polypeptid mit der Ausnahme, daß wenn es sich bei dem N-terminalen Teil von HSA um den Teil 1-n mit  $n = 369$  bis 419 oder eine Variante davon handelt, das Polypeptid aus
  - (a) dem Teil 585 bis 1578 von Humanfibronectin oder einer Variante davon,
  - 10 (b) dem Teil 1 bis 368 von CD4 oder einer Variante davon,
  - (c) dem "Platelet Derived Growth Factor" (PDGF) oder einer Variante davon,
  - (d) dem "Transforming Growth Factor  $\beta$ " (TGF  $\beta$ ) oder einer Variante davon,
  - (e) dem Teil 1-261 von reifem Humanplasmafibronektin oder einer Variante davon,
  - (f) dem Teil 278-578 von reifem Humanplasmafibronektin oder einer Variante davon,
  - 15 (g) dem Teil 1-272 von reifem Human-von Willebrand's-Faktor oder einer Variante davon oder
  - (h) Alpha-1-Antitrypsin oder einer Variante davon, besteht.
2. Fusionspolypeptid nach Anspruch 1, zusätzlich umfassend mindestens eine N-terminale Aminosäure, die länger als der dem N-terminalen Teil von HSA entsprechende Teil ist.
- 20 3. Fusionspolypeptid nach Anspruch 1 oder 2, bei dem sich an der Verbindung der N-terminalen oder C-terminalen Teile eine spaltbare Region befindet.
4. Fusionspolypeptid nach einem der vorhergehenden Ansprüche, wobei der C-terminale Teil aus dem Teil 585 bis 1578 von Humanplasmafibronektin oder einer Variante davon besteht.
- 25 5. Transformierter oder transfizierter Wirt mit einer Nukleotidsequenz, die so angeordnet ist, daß sie ein Fusionspolypeptid nach einem der vorhergehenden Ansprüche exprimieren kann.
6. Verfahren zur Herstellung eines Fusionspolypeptids durch Kultivieren eines Wirts nach Anspruch 5 und anschließendes Abtrennen des Fusionspolypeptids in einer geeigneten Form.
- 30 7. Fusionspolypeptid nach einem der Ansprüche 1 bis 4 zur therapeutischen Verwendung.

**Patentansprüche für folgende Vertragsstaaten : ES, GR**

- 35 1. Verfahren zur Herstellung eines Fusionspolypeptids durch
  - (i) Kultivieren eines transformierten oder transfizierten Wirts mit einer Nukleotidsequenz, die so angeordnet ist, daß sie ein Fusionspolypeptid exprimiert, und
  - (ii) anschließendes Abtrennen des Fusionspolypeptids in einer geeigneten Form,
  - 40 dadurch gekennzeichnet, daß das Fusionspolypeptid als mindestens einen Teil seines N-terminalen Teils einen N-terminalen Teil von HSA oder eine Variante davon und als mindestens einen Teil seines C-terminalen Teils ein weiteres Polypeptid umfaßt, mit der Ausnahme, daß wenn es sich bei dem N-terminalen Teil von HSA um den Teil 1-n mit  $n = 369$  bis 419 oder eine Variante davon handelt, das Polypeptid aus
    - (a) dem Teil 585-1578 von Humanfibronectin oder einer Variante davon,
    - 45 (b) dem Teil 1-368 von CD4 oder einer Variante davon,
    - (c) dem Platelet Derived Growth Factor oder einer Variante davon,
    - (d) dem Transforming Growth Factor  $\beta$  oder einer Variante davon,
    - (e) dem Teil 1-261 von reifem Humanplasmafibronektin oder einer Variante davon,
    - (f) dem Teil 278-578 von reifem Humanplasmafibronektin oder einer Variante davon,
    - 50 (g) dem Teil 1-272 von reifem Human-von Willebrand's-Faktor oder einer Variante davon oder
    - (h)  $\alpha$ -1-Antitrypsin oder einer Variante davon besteht.
2. Verfahren nach Anspruch 1, wobei das Fusionspolypeptid zusätzlich mindestens eine N-terminale Aminosäure, die länger als der dem N-terminalen Teil von HSA entsprechende Teil ist, umfaßt.
- 55 3. Verfahren nach Anspruch 1 oder 2, wobei sich in dem Fusionspolypeptid an der Verbindung der N-terminalen oder C-terminalen Teile eine spaltbare Region befindet.

4. Verfahren nach einem der vorhergehenden Ansprüche, wobei der C-terminale Teil aus dem Teil 585-1578 von Humanplasmafibronektin oder einer Variante davon besteht.

## 5 Revendications

**Revendications pour les Etats contractants suivants : AT, BE, CH, DE, DK, FR, IT, LU, NL, SE**

- 10 1. Polypeptide fusionné comprenant en tant qu'au moins une partie de sa portion N-terminale, une portion N-terminale de HSA ou d'un variant de celle-ci et, en tant qu'au moins une partie de sa portion C-terminale, un autre polypeptide sauf que, lorsque cette portion N-terminale de HSA est la portion 1-n dans laquelle n est 369 à 419 ou un variant de celle-ci, ce polypeptide est (a) la portion 585 à 1578 de la fibronectine humaine ou un variant de celle-ci, (b) la portion 1 à 368 de CD4 ou un variant de celle-ci, (c) le facteur de croissance dérivé des plaquettes sanguines ou un variant de celui-ci, (d) le facteur de croissance  $\beta$  de transformation ou un variant de celui-ci, (e) la portion 1-261 de la fibronectine mature de plasma humain ou un variant de celle-ci, (f) la portion 278-578 de la fibronectine mature de plasma humain ou un variant de celle-ci, (g) la portion 1-272 du facteur humain mature de von Willebrand ou un variant de celle-ci, ou (h) l'alpha-1-antitrypsine ou un variant de celle-ci.
- 20 2. Polypeptide fusionné suivant la revendication 1, comprenant de plus au moins un acide aminé N-terminal se prolongeant au-delà de la portion correspondant à la portion N-terminale de HSA.
3. Polypeptide fusionné suivant les revendications 1 ou 2, dans lequel il y a une région susceptible d'être coupée à la jonction de ces portions N-terminale et C-terminale.
- 25 4. Polypeptide fusionné suivant l'une quelconque des revendications précédentes, dans lequel cette portion C-terminale est la portion 585 à 1578 de la fibronectine de plasma humain ou un variant de celle-ci.
5. Hôte transformé ou transfecté ayant une séquence de nucléotides arrangée de façon à exprimer un polypeptide fusionné suivant l'une quelconque des revendications précédentes.
- 30 6. Procédé pour préparer un polypeptide fusionné par culture d'un hôte suivant la revendication 5, suivie de la séparation du polypeptide fusionné sous une forme utile.
7. Polypeptide fusionné suivant l'une quelconque des revendications 1 à 4 utilisable en thérapie.

**Revendications pour les Etats contractants suivants : ES, GR**

- 40 1. Procédé pour préparer un polypeptide fusionné par (i) la culture d'un hôte transformé ou transfecté ayant une séquence de nucléotides arrangée de façon à exprimer un polypeptide fusionné, suivie de (ii) la séparation du polypeptide fusionné sous une forme utile, caractérisé en ce que le polypeptide fusionné comprend, en tant qu'au moins une partie de sa portion N-terminale, une portion N-terminale de HSA ou d'un variant de celle-ci et, en tant qu'au moins une partie de sa portion C-terminale, un autre polypeptide sauf que, lorsque cette portion N-terminale de HSA est la portion 1-n dans laquelle n est 369 à 419 ou un variant de celle-ci, ce polypeptide est alors (a) la portion 585 à 1578 de la fibronectine humaine ou un variant de celle-ci, (b) la portion 1 à 368 de CD4 ou un variant de celle-ci, (c) le facteur de croissance dérivé des plaquettes sanguines ou un variant de celui-ci, (d) le facteur de croissance  $\beta$  de transformation ou un variant de celui-ci, (e) la portion 1-261 de la fibronectine mature de plasma humain ou un variant de celle-ci, (f) la portion 278-578 de la fibronectine mature de plasma humain ou un variant de celle-ci, (g) la portion 1-272 du facteur humain mature de von Willebrand ou un variant de celle-ci, ou (h) l'alpha-1-antitrypsine ou un variant de celle-ci.
- 50 2. Procédé suivant la revendication 1, dans lequel le polypeptide fusionné comprend de plus au moins un acide aminé N-terminal se prolongeant au-delà de la portion correspondant à la portion N-terminale de HSA.
- 55 3. Procédé suivant les revendications 1 ou 2 dans lequel, dans le polypeptide fusionné, il y a une région susceptible d'être coupée à la jonction de ces portions N-terminale et C-terminale.

4. Procédé suivant l'une quelconque des revendications précédentes, dans lequel cette portion C-terminale est la portion 585 à 1578 de la fibronectine de plasma humain ou un variant de celle-ci.

5

10

15

20

25

30

35

40

45

50

55

FIGURE 1

10	20
Asp Ala His Lys Ser Glu Val Ala His Arg Phe Lys Asp Leu Gly Glu Glu Asn Phe Lys	
30	40
Ala Leu Val Leu Ile Ala Phe Ala Gln Tyr Leu Gln Gln Cys Pro Phe Glu Asp His Val	
50	60
Lys Leu Val Asn Glu Val Thr Glu Phe Ala Lys Thr Cys Val Ala Asp Glu Ser Ala Glu	
70	80
Asn Cys Asp Lys Ser Leu His Thr Leu Phe Gly Asp Lys Leu Cys Thr Val Ala Thr Leu	
90	100
Arg Glu Thr Tyr Gly Glu Met Ala Asp Cys Cys Ala Lys Gln Glu Pro Glu Arg Asn Glu	
110	120
Cys Phe Leu Gln His Lys Asp Asp Asn Pro Asn Leu Pro Arg Leu Val Arg Pro Glu Val	
130	140
Asp Val Met Cys Thr Ala Phe His Asp Asn Glu Glu Thr Phe Leu Lys Lys Tyr Leu Tyr	
150	160
Glu Ile Ala Arg Arg His Pro Tyr Phe Tyr Ala Pro Glu Leu Leu Phe Phe Ala Lys Arg	
170	180
Tyr Lys Ala Ala Phe Thr Glu Cys Cys Gln Ala Ala Asp Lys Ala Ala Cys Leu Leu Pro	
190	200
Lys Leu Asp Glu Leu Arg Asp Glu Gly Lys Ala Ser Ser Ala Lys Gln Arg Leu Lys Cys	
210	220
Ala Ser Leu Gln Lys Phe Gly Glu Arg Ala Phe Lys Ala Trp Ala Val Ala Arg Leu Ser	
230	240
Gln Arg Phe Pro Lys Ala Glu Phe Ala Glu Val Ser Lys Leu Val Thr Asp Leu Thr Lys	
250	260
Val His Thr Glu Cys Cys His Gly Asp Leu Leu Glu Cys Ala Asp Asp Arg Ala Asp Leu	
270	280
Ala Lys Tyr Ile Cys Glu Asn Gln Asp Ser Ile Ser Ser Lys Leu Lys Glu Cys Cys Glu	
290	300
Lys Pro Leu Leu Glu Lys Ser His Cys Ile Ala Glu Val Glu Asn Asp Glu Met Pro Ala	
310	320
Asp Leu Pro Ser Leu Ala Ala Asp Phe Val Glu Ser Lys Asp Val Cys Lys Asn Tyr Ala	
330	340
Glu Ala Lys Asp Val Phe Leu Gly Met Phe Leu Tyr Glu Tyr Ala Arg Arg His Pro Asp	
350	360
Tyr Ser Val Val Leu Leu Leu Arg Leu Ala Lys Thr Tyr Glu Thr Thr Leu Glu Lys Cys	
370	380
Cys Ala Ala Ala Asp Pro His Glu Cys Tyr Ala Lys Val Phe Asp Glu Phe Lys Pro Leu	

FIGURE 1 Cont.

	390	400
Val Glu Glu Pro Gln Asn Leu Ile Lys Gln Asn Cys Glu Leu Phe Glu Gln Leu Gly Glu		
	410	420
Tyr Lys Phe Gln Asn Ala Leu Leu Val Arg Tyr Thr Lys Lys Val Pro Gln Val Ser Thr		
	430	440
Pro Thr Leu Val Glu Val Ser Arg Asn Leu Gly Lys Val Gly Ser Lys Cys Cys Lys His		
	450	460
Pro Glu Ala Lys Arg Met Pro Cys Ala Glu Asp Tyr Leu Ser Val Val Leu Asn Gln Leu		
	470	480
Cys Val Leu His Glu Lys Thr Pro Val Ser Asp Arg Val Thr Lys Cys Cys Thr Glu Ser		
	490	500
Leu Val Asn Arg Arg Pro Cys Phe Ser Ala Leu Glu Val Asp Glu Thr Tyr Val Pro Lys		
	510	520
Glu Phe Asn Ala Glu Thr Phe Thr Phe His Ala Asp Ile Cys Thr Leu Ser Glu Lys Glu		
	530	540
Arg Gln Ile Lys Lys Gln Thr Ala Leu Val Glu Leu Val Lys His Lys Pro Lys Ala Thr		
	550	560
Lys Glu Gln Leu Lys Ala Val Met Asp Asp Phe Ala Ala Phe Val Glu Lys Cys Cys Lys		
	570	580
Ala Asp Asp Lys Glu Thr Cys Phe Ala Glu Glu Gly Lys Lys Leu Val Ala Ala Ser Gln		
Ala Ala Leu Gly Leu		

FIGURE 2 DNA sequence coding for mature HSA

```

      10      20      30      40      50      60      70      80
GATGCACACAAGAGTGAGGTTGCTCATCGGTTTAAAGATTGGGAGAAGAAAATTTCAAAGCCTTGGTGTGATTGCCTT
D A H K S E V A H R F K D L G E E N F K A L V L I A F

      90     100     110     120     130     140     150     160
TGCTCAGTATCTTCAGCAGTGTCCATTTGAAGATCATGTAAAATTAGTGAATGAAGTAACTGAATTTGCAAAAACATGTG
A Q Y L Q Q C F F E D H V K L V N E V T E F A K T C

      170     180     190     200     210     220     230     240
TTGCTGATGAGTCAGCTGAAAATTGTGACAAATCACTTCATACCCCTTTTGGAGACAAAATTATGCACAGTTGCAACTCTT
V A D E S A E N C D K S L H T L F G D K L C T V A T L

      250     260     270     280     290     300     310     320
CGTGAACCTATGGTGAATGGCTGACTGCTGTGCAAAACAAGAACCTGAGAGAAATGAATGCTTCTTGCACACAAAAGA
R E T Y G E M A D C C A K Q E P E R N E C F L Q H K D

      330     340     350     360     370     380     390     400
TGACAACCCAAACCTCCCCGATTGGTGAGACCAGAGGTTGATGTGATGTGCACTGCTTTTCATGACAATGAAGAGACAT
D N P N L P R L V R P E V D V M C T A F H D N E E T

      410     420     430     440     450     460     470     480
TTTTGAAAAATACTTATATGAATTGCCAGAAGACATCCTTACTTTTATGCCCCGGAACTCCTTTTCTTGTCTAAAAGG
F L K K Y L Y E I A R R H P Y F Y A P E L L F F A K R

      490     500     510     520     530     540     550     560
TATAAAGCTGCTTTTACAGAATGTTGCCAAGCTGCTGATAAAGCTGCCCTGCTTGGCCAAAGCTCGATGAACCTCGGGGA
Y K A A F T E C C Q A A D K A A C L L P K L D E L R D

      570     580     590     600     610     620     630     640
TGAAGGGGAAGGCTTCGTCTGCCAAACAGAGACTCAAATGTGCCAGTCTCCAAAAATTTGGAGAAAGAGCTTTCAAAGCAT
E G K A S S A K Q R L K C A S L Q K F G E R A F K A

      650     660     670     680     690     700     710     720
GGGCAGTGGCTCGCCTGAGCCAGAGATTTCCCAAAGCTGAGTTTGCAGAAGTTTCCAAGTTAGTGACAGATCTTACCAAA
W A V A R L S Q R F P K A E F A E V S K L V T D L T K

      730     740     750     760     770     780     790     800
GTCCACACGGAATGCTGCCATGGAGATCTGCTTGAATGTGCTGATGACAGGGCGGACCTTGCCAAGTATATCTGTGAAA
V H T E C C H G D L L E C A D D R A D L A K Y I C E N

      810     820     830     840     850     860     870     880
TCAGGATTCGATCTCCAGTAAACTGAAGGAATGCTGTGAAAAACCTCTGTTGGAAAAATCCCACTGCATTGCCGAAGTGG
Q D S I S S K L K E C C E K P L L E K S H C I A E V

      890     900     910     920     930     940     950     960
AAAATGATGAGATGCCTGCTGACTTGCCTTCATTAGCTGCTGATTTTGTGAAAGTAAGGATGTTTGCAAAAACTATGCT
E N D E M P A D L P S L A A D F V E S K D V C K N Y A

      970     980     990    1000    1010    1020    1030    1040
GAGGCAAGGATGTCTTCTGGGCATGTTTTTGTATGAATATGCAAGAAGGCATCCTGATTACTCTGTCGTGCTGCTGCT
E A K D V F L G M F L Y E Y A R R H P D Y S V V L L L

```

FIGURE 2 Cont.

1050 1060 1070 1080 1090 1100 1110 1120  
 GAGACTTGCCAAGACATATGAAACCACCTCTAGAGAAGTGCTGTGCCCGCTGCAGATCCTCATGAATGCTATGCCAAAGTGT  
 R L A K T Y E T T L E X C C A A A D P H E C Y A K V

1130 1140 1150 1160 1170 1180 1190 1200  
 TCGATGAATTTAAACCTCTGTGGAGAGCCCTCAGAATTTAATCAAACAAAAGTGTGAGCTTTTGGAGCAGCTTGGAGAG  
 F D E F K P L V E E P Q N L L K Q N C E L F E Q L G E

1210 1220 1230 1240 1250 1260 1270 1280  
 TACAAATTCAGAATGCCTATTAGTTCTGTACACCAAGAAAGTACCCCAAGTGTCAACTCCAACCTCTGTAGAGGTCTC  
 Y K F Q N A L L V R Y T K K V P Q V S T P T L V E V S

1290 1300 1310 1320 1330 1340 1350 1360  
 AAGAAACCTAGGAAAAGTGGGACGCAATGTTGTAACATCCTGAAGCAAAAAGAAATGCCCTGTGCAGAGACTATCTAT  
 R N L G K V G S K C C X H P E A K R M P C A E D Y L

1370 1380 1390 1400 1410 1420 1430 1440  
 CCGTGGTCTGAACCAAGTATGTGTGTGCATGAGAAAACGCCAGTAAGTGACAGAGTCACAAAATGCTGCACAGAGTCC  
 S V V L N Q L C V L H E K T P V S D R V T K C C T E S

1450 1460 1470 1480 1490 1500 1510 1520  
 TTGGTGAACAGGCGACCATGCTTTTCAGCTCTGGAAAGTCGATGAAACATACGTTCCCAAAGAGTTTAATGCTGAAACATT  
 L V N R R P C F S A L E V D E T Y V P K E F N A E T F

1530 1540 1550 1560 1570 1580 1590 1600  
 CACCTTCCATGCAGATATATGCACACTTTCTGAGAAGGAGAGACAAATCAAGAAACAAACTGCCTTGTGTGAGCTTGTGA  
 T F H A D E C T L S E K E R Q I K K Q T A L V E L V

1610 1620 1630 1640 1650 1660 1670 1680  
 AACACAAGCCCAAGGCAACAAAAGAGCAACTGAAAGCTGTTATGGATGATTTCCGAGCTTTTGTAGAGAAGTGCTGCAAG  
 K H K P K A T K E Q L K A V M D D F A A F V E K C C K

1690 1700 1710 1720 1730 1740 1750 1760  
 GCTGACGATAAGGAGACCTGCTTTGCCGAGGAGGGTAAAAAAGTGTGCTGCAAGTCAAGCTGCCCTTAGGCTTATAACA  
 A D D K E T C F A E E G K K L V A A S Q A A L G L

1770 1780  
 TCTACATTTAAAAAGCATCTCAG

FIGURE 3 Construction of mEOB16

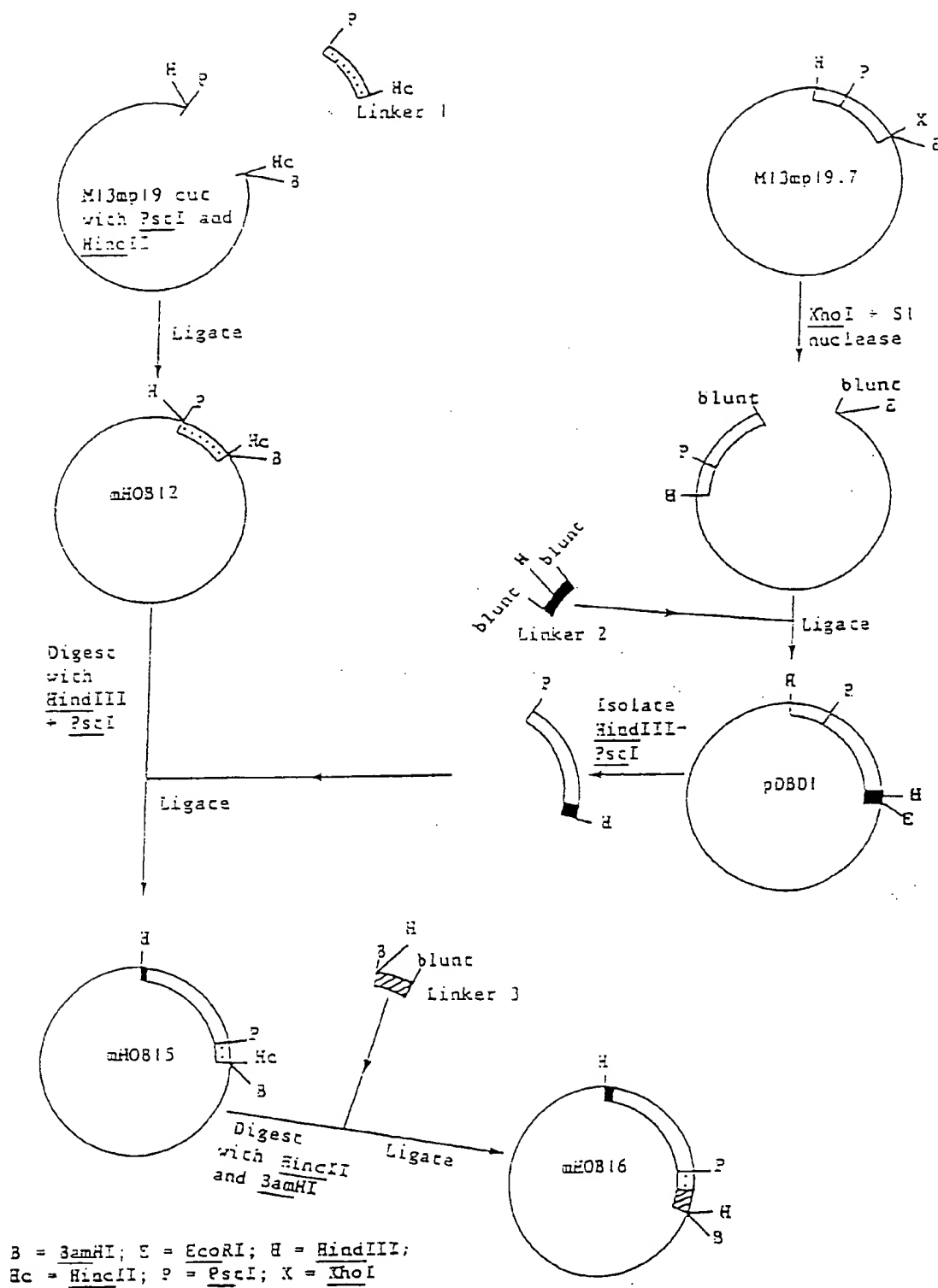
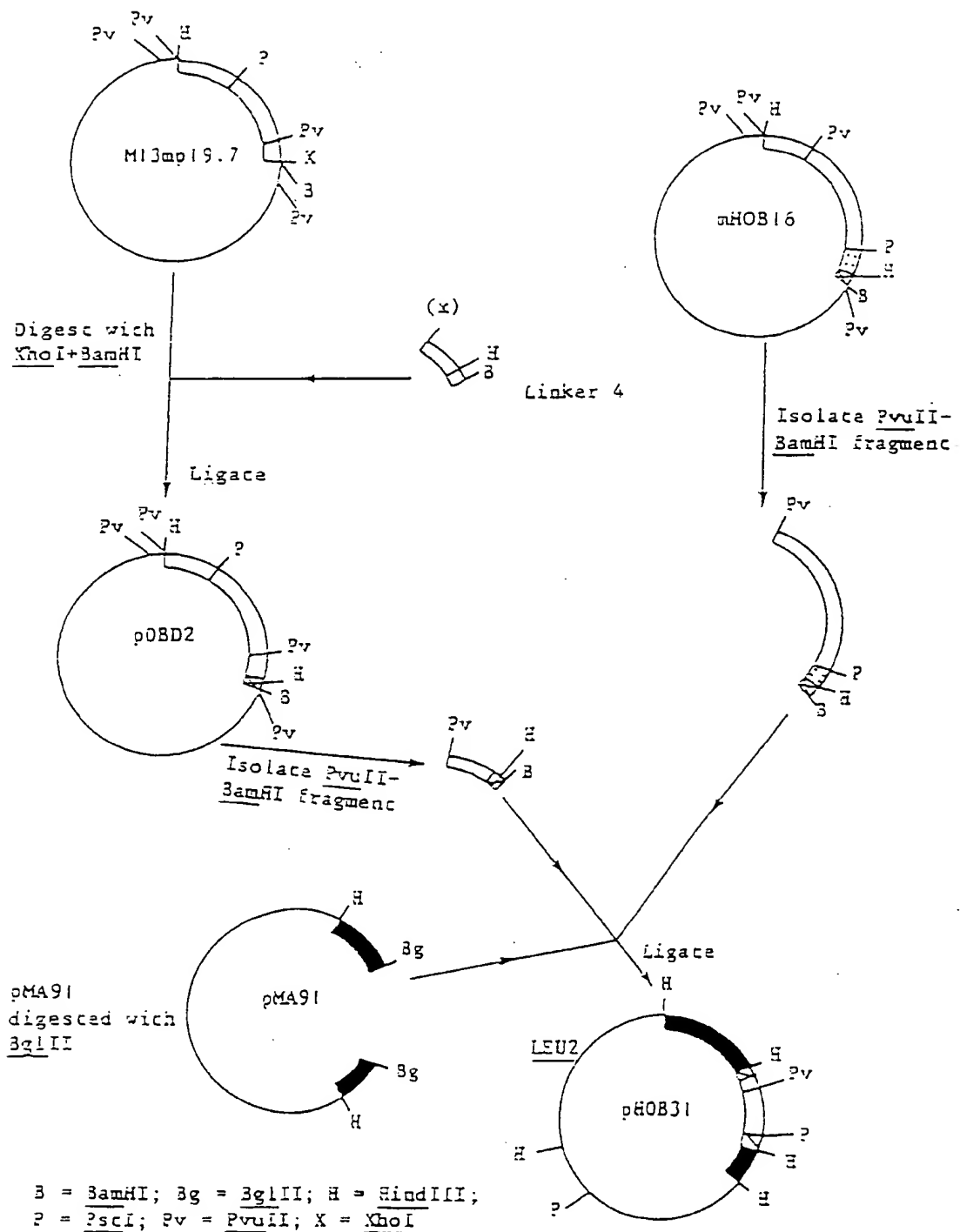


FIGURE 4 Construction of pHOB31



Gln Ala Gln Gln Met Val Gln Pro Gln Ser Pro Val Ala Val Ser Gln Ser Lys Pro Gly  
 Cys Tyr Asp Asn Gly Lys His Tyr Gln Ile Asn Gln Gln Trp Glu Arg Thr Tyr Leu Gly  
 Asn Val Leu Val Cys Thr Cys Tyr Gly Gly Ser Arg Gly Phe Asn Cys Glu Ser Lys Pro  
 Glu Ala Glu Glu Thr Cys Phe Asp Lys Tyr Thr Gly Asn Thr Tyr Arg Val Gly Asp Thr  
 Tyr Glu Arg Pro Lys Asp Ser Met Ile Trp Asp Cys Thr Cys Ile Gly Ala Gly Arg Gly  
 Arg Ile Ser Cys Thr Ile Ala Asn Arg Cys His Glu Gly Gly Gln Ser Tyr Lys Ile Gly  
 Asp Thr Trp Arg Arg Pro His Glu Thr Gly Gly Tyr Met Leu Glu Cys Val Cys Leu Gly  
 Asn Gly Lys Gly Glu Trp Thr Cys Lys Pro Ile Ala Glu Lys Cys Phe Asp His Ala  
 Gly Thr Ser Tyr Val Val Gly Glu Thr Trp Glu Lys Pro Tyr Gln Gly Trp Met Met Val  
 Asp Cys Thr Cys Leu Gly Glu Gly Ser Arg Ile Thr Cys Thr Ser Arg Asn Arg Cys  
 Asn Asp Gln Asp Thr Arg Thr Ser Tyr Arg Ile Gly Asp Thr Trp Ser Lys Lys Asp Asn  
 Arg Gly Asn Leu Leu Gln Cys Ile Cys Thr Gly Asn Gly Arg Gly Glu Trp Lys Cys Glu  
 Arg His Thr Ser Val Gln Thr Thr Ser Ser Gly Ser Gly Pro Phe Thr Asp Val Arg Ala  
 Ala Val Tyr Gln Pro Gln Pro His Pro Gln Pro Pro Tyr Gly His Cys Val Thr Asp  
 Ser Ely Val Val Tyr Ser Val Gly Met Gln Trp Leu Lys Thr Gln Gly Asn Lys Gln Met  
 Leu Cys Thr Cys Leu Gly Asn Gly Val Ser Cys Gln Glu Thr Ala Val Thr Gln Thr  
 Gly Gly Asn Ser Asn Gly Glu Pro Cys Val Leu Pro Phe Thr Tyr Asn Gly Arg Thr  
 Tyr Ser Cys Thr Thr Glu Gly Arg Gln Asp Gly His Leu Trp Cys Ser Thr Thr Ser Asn  
 Tyr Glu Gln Asp Gln Lys Tyr Ser Phe Cys Thr Asp His Thr Val Leu Val Gln Thr  
 Gly Gly Asn Ser Asn Ely Ala Leu Cys His Phe Pro Phe Leu Tyr Asn Asn His Asn Tyr  
 Thr Asp Cys Thr Ser Glu Gly Arg Arg Asp Asn Met Lys Trp Cys Gly Thr Thr Gln Asn

Fig. 5A

Tyr Asp Ala Asp Gln Lys Phe Gly Phe Gly Cys Pro Met Ala Ala His Glu Glu Ile Cys 430 Thr  
 Thr Asn Glu Gly Val Met Tyr Arg Ile Gly Asp Gln Trp Asp Lys Gln His Asp Met 440 Gly  
 His Met Met Arg Cys Thr Cys Val Gly Asn Gly Arg Gly Glu Trp Thr Cys Tyr Ala 450 Tyr  
 Ser Gln Leu Arg Asp Gln Cys Ile Val Asp Asp Ile Thr Tyr Asn Val Asn Asp Thr 460 Phe  
 His Lys Arg His Glu Glu Gly His Met Leu Asn Cys Thr Cys Phe Gly Gln Gly Arg 470 Gly  
 Arg Trp Lys Cys Asp Pro Val Asp Gln Cys Gln Asp Ser Glu Thr Gly Thr Phe Tyr 480 Gln  
 Ile Gly Asp Ser Trp Glu Lys Tyr Val His Gly Val Arg Tyr Gln Cys Tyr Cys Tyr 490 Gly  
 Arg Gly Ile Gly Glu Trp His Cys Gln Pro Leu Gln Thr Tyr Pro Ser Ser Ser Gly 500 Pro  
 Val Glu Val Phe Ile Thr Glu Thr Pro Ser Gln Pro Asn Ser His Pro Ile Gln Trp Asn 510  
 Ala Pro Gln Pro Ser His Ile Ser Lys Tyr Ile Leu Arg Trp Arg Pro Lys Asn Ser 520 Val  
 Gly Arg Trp Lys Glu Ala Thr Ile Pro Gly His Leu Asn Ser Tyr Thr Ile Lys Gly 530 Leu  
 Lys Pro Gly Val Val Tyr Glu Gly Gln Leu Ile Ser Ile Gln Gln Tyr Gly His Gln 540 Glu  
 Val Thr Arg Phe Asp Phe Thr Thr Thr Ser Thr Pro Val Thr Ser Asn Thr 550 Val  
 Thr Gly Glu Thr Thr Pro Phe Ser Pro Leu Val Ala Thr Ser Glu Ser Val Thr Glu 560 Ile  
 Thr Ala Ser Ser Phe Val Val Ser Trp Val Ser Ala Ser Asp Thr Val Ser Gly Phe Arg 570  
 Val Glu Tyr Glu Leu Ser Glu Glu Gly Asp Glu Pro Gln Tyr Leu Asp Leu Pro Ser 580 Thr  
 Ala Thr Ser Val Asn Ile Pro Asp Leu Leu Pro Ely Arg Lys Tyr Ile Val Asn Val 590  
 Gln Ile Ser Glu Asp Gly Glu Gln Ser Leu Ile Leu Ser Thr Ser Gln Thr Thr Ala 600 Pro  
 Asp Ala Pro Pro Asp Pro Thr Val Asp Gln Val Asp Asp Thr Ser Ile Val Val Arg Trp 610  
 Ser Arg Pro Gln Ala Pro Ile Thr Gly Tyr Arg Ile Val Tyr Ser Pro Ser Val Glu 620 Gly  
 Ser Ser Thr Glu Leu Asn Leu Pro Glu Thr Ala Asn Ser Val Thr Leu Ser Asp Leu Gln 630  
 FNDEL 1

Fig. 5B

Pro Gly Val Gln Tyr Asn Ile Thr Ile Tyr Ala Val Glu Glu Asn Gln Glu Ser Thr 850 860  
 Val Val Ile Gln Gln Thr Thr Gly Thr Pro Arg Ser Asp Thr Val Pro Ser Pro Arg 870 880  
 Aso Leu Gln Phe Val Glu Val Thr Asp Val Lys Val Thr Ile Met Trp Thr Pro Pro Glu 890 900  
 Ser Ala Val Thr Gly Tyr Arg Val Asp Val Ile Pro Val Asn Leu Pro Gly Glu His Gly 910 920  
 Gln Arg Leu Pro Ile Ser Arg Asn Thr Phe Ala Glu Val Thr Gly Leu Ser Pro Gly Val 930 940  
 Thr Tyr Tyr Phe Lys Val Phe Ala Val Ser His Gly Arg Glu Ser Lys Pro Leu Thr Ala 950 960  
 Gln Gln Thr Thr Lys Leu Asp Ala Pro Thr Asn Leu Gln Phe Val Asn Glu Thr Asp Ser 970 980  
 Thr Val Leu Val Arg Trp Thr Pro Pro Arg Ala Gln Ile Thr Gly Tyr Arg Leu Thr Val 990 1000  
 Gly Leu Thr Arg Arg Gly Gln Pro Arg Gln Tyr Asn Val Gly Pro Ser Val Ser Lys Tyr 1010 1020  
 Pro Leu Arg Asn Leu Gln Pro Ala Ser Glu Tyr Thr Val Ser Leu Val Ala Ile Lys Gly 1030 1040  
 Asn Gln Glu Ser Pro Lys Ala Thr Gly Val Phe Thr Thr Leu Gln Pro Gly Ser Ser Ile 1050 1060  
 Pro Pro Tyr Asn Thr Glu Val Thr Glu Thr Thr Ile Val Ile Thr Trp Thr Pro Ala Pro 1070 1080  
 Arg Ile Gly Phe Lys Leu Gly Val Arg Pro Ser Gln Gly Gly Glu Ala Pro Arg Glu Val 1090 1100  
 Thr Ser Asp Ser Gly Ser Ile Val Val Ser Gly Leu Thr Pro Gly Val Glu Tyr Val Tyr 1110 1120  
 Thr Ile Gln Val Leu Arg Asp Gly Gln Glu Arg Asp Ala Pro Ile Val Asn Lys Val Val 1130 1140  
 Thr Pro Leu Ser Pro Pro Thr Asn Leu His Leu Glu Ala Asn Pro Asp Thr Gly Val Leu 1150 1160  
 Thr Val Ser Trp Glu Arg Ser Thr Thr Pro Asp Ile Thr Gly Tyr Arg Ile Thr Thr Thr 1170 1180  
 Pro Thr Asn Gly Gln Gln Gly Asn Ser Leu Glu Glu Val Val His Ala Asp Gln Ser Ser 1190 1200  
 Cys Thr Phe Asp Asn Leu Ser Pro Gly Leu Glu Tyr Asn Val Ser Val Tyr Thr Val Lys 1210 1220  
 Asp Asp Lys Glu Ser Val Pro Ile Ser Asp Thr Ile Ile Pro Ala Val Pro Pro Thr 1230 1240  
 Asp Leu Arg Phe Thr Asn Ile Gly Pro Asp Thr Met Arg Val Thr Trp Ala Pro Pro Pro 1250 1260

Fig. 5C

Ser Ile Asp Leu Thr Asn Phe Leu Val Arg Tyr Ser Pro Val Lys Asn Glu Glu Asp Val  
 Ala Glu Leu Ser Ile Ser Pro Ser Asp Asn Ala Val Val Leu Thr Asn Leu Leu Pro Gly  
 Thr Glu Tyr Val Val Ser Val Ser Val Tyr Glu Gln His Glu Ser Thr Pro Leu Arg  
 Gly Arg Gln Lys Thr Gly Leu Asp Ser Thr Gly Ile Asp Phe Ser Asp Ile Thr Ala  
 Asn Ser Phe Thr Val His Trp Ile Ala Pro Arg Ala Thr Ile Thr Gly Tyr Arg Ile Arg  
 His His Pro Glu His Phe Ser Gly Arg Pro Arg Glu Asp Arg Val Pro His Ser Arg Asn  
 Ser Ile Thr Leu Thr Asn Lau Thr Pro Gly Thr Glu Tyr Val Val Ser Ile Val Ala Leu  
 Asn Gly Arg Glu Glu Ser Pro Leu Leu Ile Gly Gln Gln Ser Thr Val Ser Asp Val Pro  
 Arg Asp Leu Glu Val Val Ala Ala Thr Pro Thr Ser Leu Leu Ile Ser Trp Asp Ala Pro  
 Ala Val Thr Val Arg Tyr Tyr Arg Ile Thr Tyr Gly Glu Thr Gly Asn Ser Pro Val  
 Gln Glu Phe Thr Val Pro Gly Ser Lys Ser Thr Ala Thr Ile Ser Gly Leu Lys Pro Gly  
 Val Asp Tyr Thr Ile Thr Val Tyr Ala Val Thr Gly Arg Gly Asp Ser Pro Ala Ser Ser  
 Lys Pro Ile Ser Ile Asn Tyr Arg Thr Glu Ile Asp Lys Pro Ser Gln Met Gln Val Thr  
 Asp Val Gln Asp Asn Ser Ile Ser Val Lys Trp Leu Pro Ser Ser Pro Val Thr Gly  
 Tyr Arg Val Thr Thr Thr Pro Lys Asn Gly Pro Gly Pro Thr Lys Thr Lys Thr Ala Gly  
 Pro Asp Gln Thr Glu Met Thr Ile Glu Gly Leu Gln Pro Thr Thr Val Glu Tyr Val Val Ser  
 Val Tyr Ala Gln Asn Pro Ser Gly Glu Ser Gln Pro Leu Val Gln Thr Ala Val Thr Thr  
 Ile Pro Ala Pro Thr Asp Leu Lys Phe Thr Gln Val Thr Pro Thr Ser Leu Ser Ala Gln  
 Trp Thr Pro Pro Asn Val Gln Leu Thr Gly Tyr Arg Val Arg Val Thr Pro Lys Glu Lys  
 Thr Gly Pro Met Lys Glu Ile Asn Leu Ala Pro Asp Ser Ser Val Val Val Ser Gly  
 Leu Met Val Ala Thr Lys Tyr Glu Val Ser Val Tyr Ala Leu Lys Asp Thr Leu Thr Ser  
 FNDEL I

Fig. 5D

24

Val	Asn	Tyr	Lys	Ile	Gly	Glu	Lys	Trp	Asp	Arg	Gln	Gly	Glu	Asn	Gly	Gln	Met	Met	Ser	2120
Cys	Thr	Cys	Leu	Gly	Asn	Gly	Lys	Gly	Glu	Phe	Lys	Cys	Asp	Pro	His	Glu	Ala	Thr	Cys	2140
Tyr	Asp	Asp	Gly	Lys	Thr	Tyr	His	Val	Gly	Glu	Gln	Trp	Gln	Lys	Glu	Tyr	Leu	Gly	Ala	2160
Ile	Cys	Ser	Cys	Thr	Cys	Phe	Gly	Gly	Gln	Arg	Gly	Trp	Arg	Cys	Asp	Asn	Cys	Arg	Arg	2180
Pro	Gly	Gly	Glu	Pro	Ser	Pro	Glu	Gly	Thr	Thr	Gly	Gln	Ser	Tyr	Asn	Gln	Tyr	Ser	Gln	2200
Arg	Tyr	His	Gln	Arg	Thr	Asn	Thr	Asn	Val	Asn	Cys	Pro	Ile	Glu	Cys	Phe	Met	Pro	Leu	2220
Asp	Val	Gln	Ala	Asp	Arg	Glu	Asp	Ser	Arg	Glu										2230

Fig. 5F

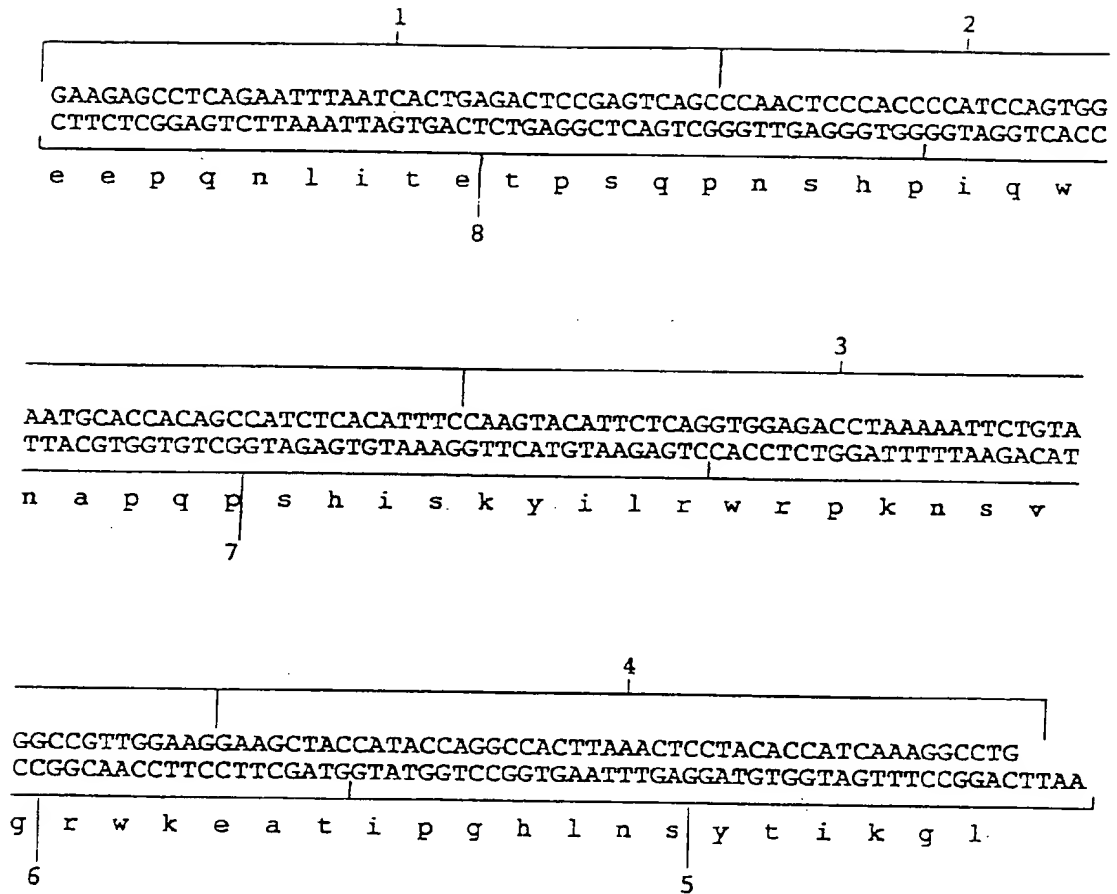


Figure 6 Linker 5 showing the eight constituent oligonucleotides

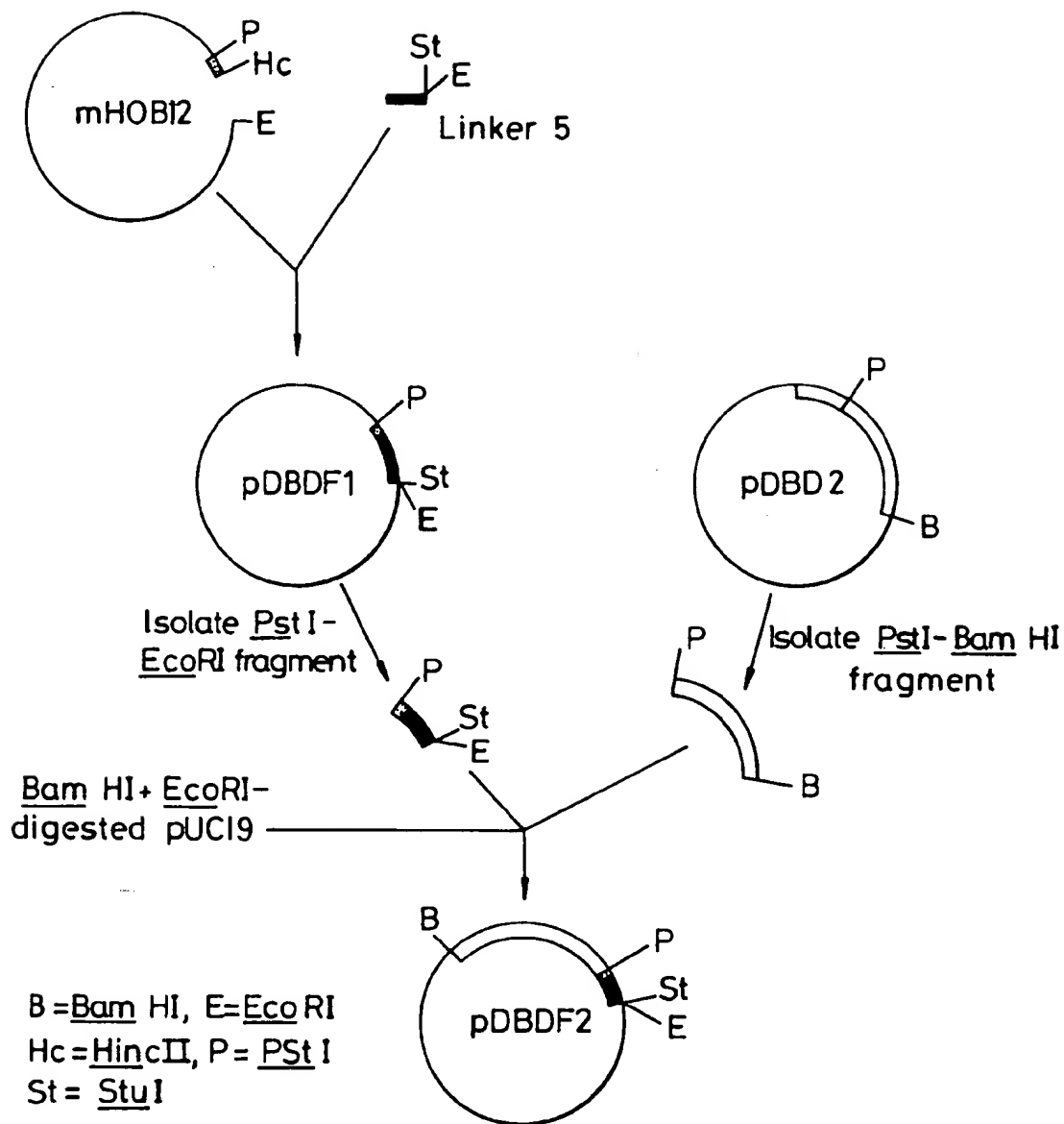


Fig. 7 Construction of pDBDF2

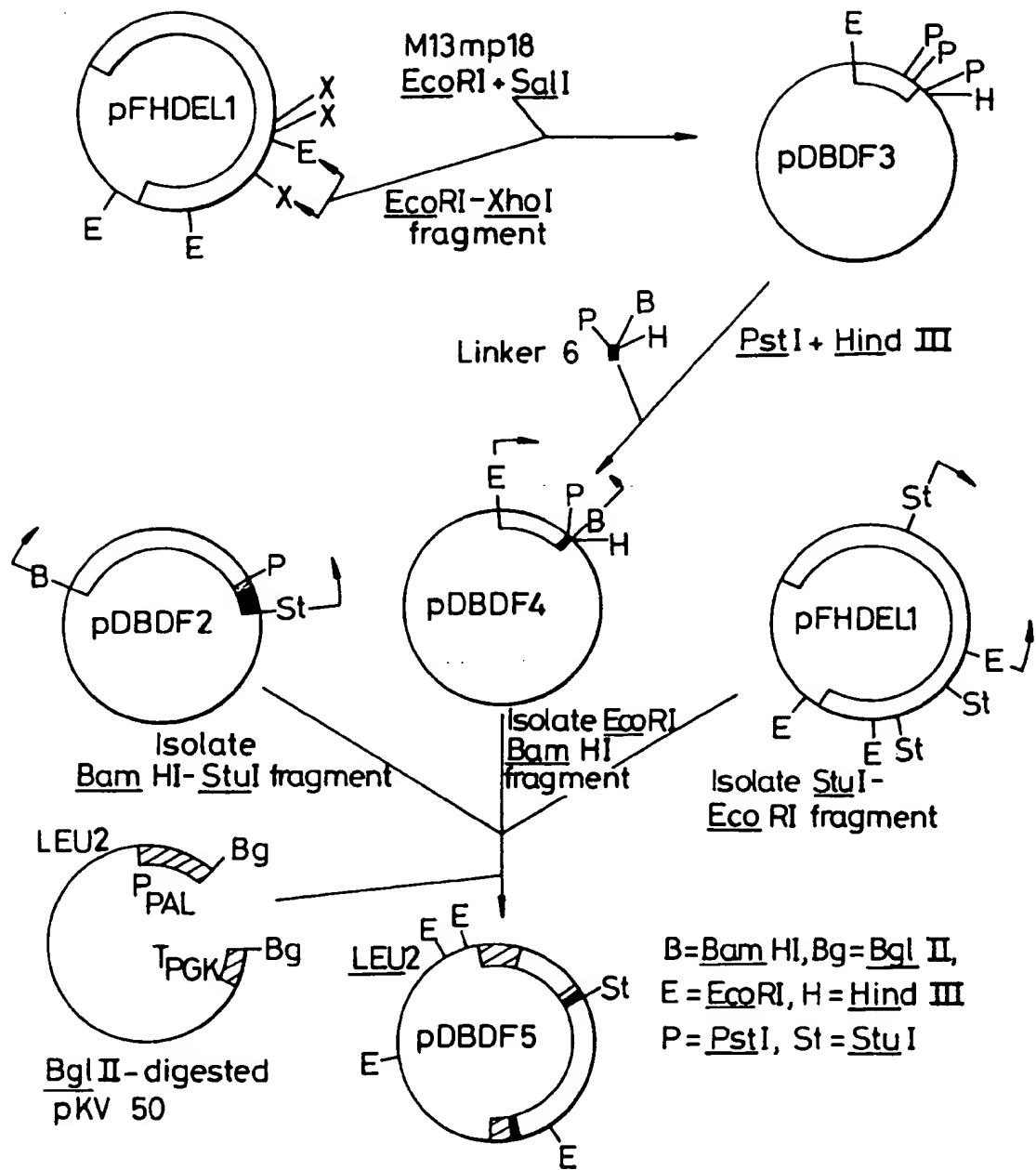


Fig. 8 Construction of pDBDF5

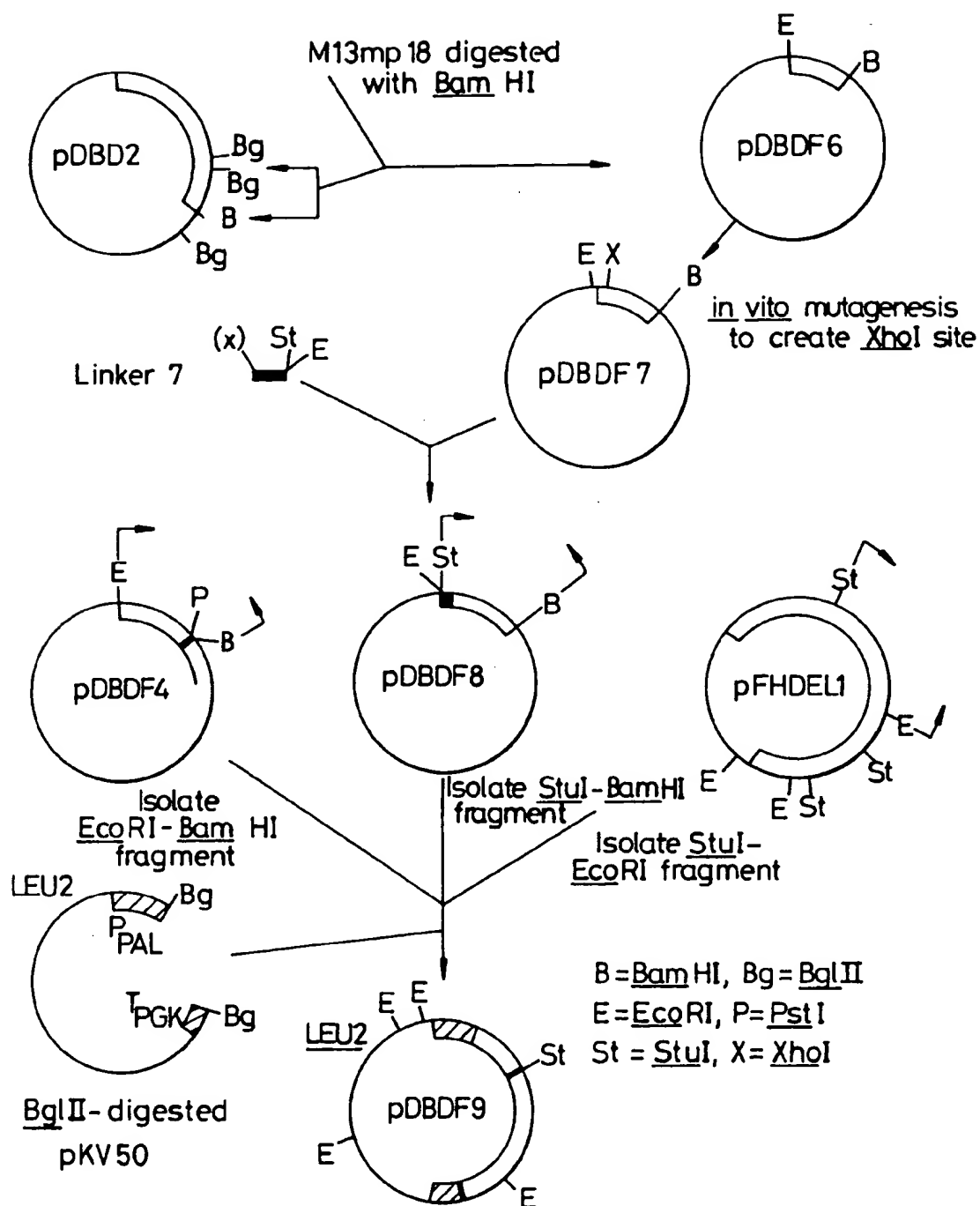


Fig. 9 Construction of pDBDF9

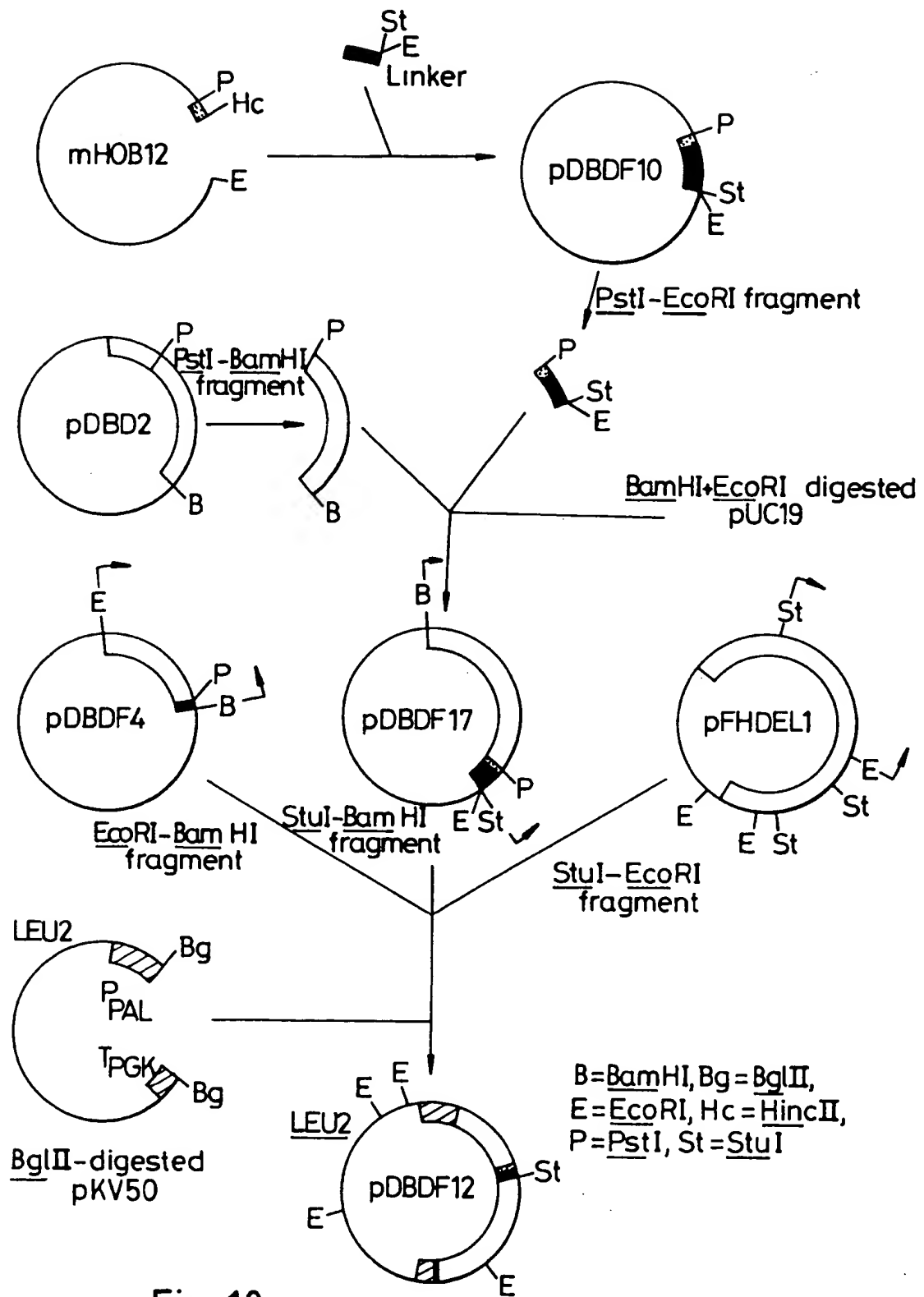


Fig. 10 Construction of pDBDF12

Figure 11

Name: pFHDEL1  
 Vector: pUC18 Amp<sup>r</sup> 2860bp  
 Insert: hFNCdNA - 7630bp

